

PARTIAL TRANSLATION OF JP 10(1998)-74095 A

Publication Date: March 17, 1998

Title of the Invention: VOICE CODING DEVICE AND VOICE DECODING
DEVICE

Patent Application Number: 8-232388

Filing Date: September 2, 1996

Applicant: SHARP CORP

Inventors: Nobuyoshi UMIKI

(Page 3, right column, lines 5-30)

[0010] The pitches of inputted voices are analyzed and extracted in an analysis interval unit. The extracted pitches are made into a model by the addition of the accents, the phrases, and each pattern of phonemes. The error between the pitch of actual voice and the patterned pitch and each pattern are transmitted, whereby compressed and coded voice code data is generated. By the addition of the accents, the phrases, each pattern of phonemes, and the error between the patterned pitch and the pitch of the actual voice, pitches are generated and decoded the voice code data.

[0011] Furthermore, it is considered that the above-mentioned problem of discontinuity of the pitch between phonemes is caused by the fact that a pitch pattern has not been made into a model. Thus, by using a model based on a voice generation mechanism for pitch control, the discontinuity of a pitch can be eliminated. If the discontinuity of a pitch is eliminated, the above-mentioned problems that the height of a voice is rapidly changed, an abnormal sound is generated, and a voice sounds unnaturally can be overcome, and more natural voice synthesis can be conducted.

[0012] Furthermore, by providing model parameter control means for controlling and altering a model parameter of modeled and compressed pitch information in accordance with a pitch model, even based on a pitch model, a

modeled parameter is controlled and altered so that a pitch is not increased to high, for example, even in the case where a large accent component is continued, whereby a pitch can be prevented from becoming too high or too low.



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **10074095 A**(43) Date of publication of application: **17 . 03 . 98**

(51) Int. Cl. **G10L 3/00**
G10L 9/00

(21) Application number: **08232388**(71) Applicant: **SHARP CORP**(22) Date of filing: **02 . 09 . 96**(72) Inventor: **UMIKI NOBUYOSHI**

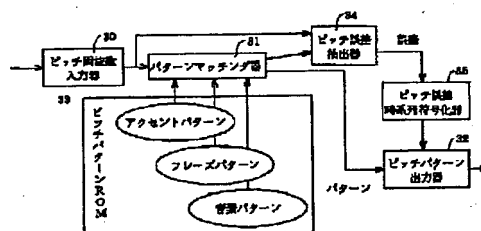
(54) **VOICE CODING DEVICE AND VOICE DECODING
 DEVICE**

(57) Abstract:

PROBLEM TO BE SOLVED: To increase the compression rate of the compression-encoding of voices and to resolve the discontinuity of the pitch while the phoneme pieces, which are compressed-encoded, are edited and synthesized voices are outputted.

SOLUTION: The pitches of inputted voices are analyzed, extracted in an analysis interval unit. The extracted pitches are then made into a model by the addition of the accents, the phrases and the patterns of phonemes stored in a pitch pattern ROM 33. Moreover, a pitch error extractor 34 extracts the error between the pitch of actual voice and the patterned pitch and generates compressed and coded voice code data by transmitting each pattern and its error. The decoder generates, decodes pitches by adding the accents, the phrases, each pattern of phonemes and the error between the patterned pitch and the pitch of the actual voice.

COPYRIGHT: (C)1998,JPO



(19) 日本国特許庁(JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平 10-74095

(43) 公開日 平成10年(1998)3月17日

(51) Int. Cl. [°]	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L	3/00		G 1 0 L	H
	9/00			J

審査請求 未請求 請求項の数 14 O L (全 9 頁)

(21) 出願番号 特願平 8-232388

(22) 出願日 平成8年(1996)9月2日

(71) 出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72) 発明者 海木 延佳

大阪府大阪市阿倍野区長池町22番22号 シ
ャープ株式会社内

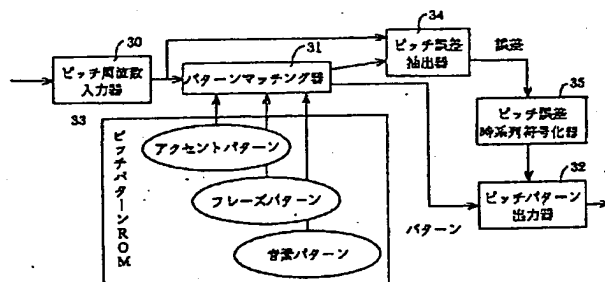
(74) 代理人 弁理士 平木 祐輔

(54) 【発明の名称】 音声符号化装置及び音声復号化装置

(57) 【要約】

【課題】 音声の圧縮符号化の圧縮率を高め、また圧縮符号化された音素片を編集して合成音声出力するときのピッチの不連続性を解消する。

【解決手段】 入力された音声のピッチを分析間隔単位に分析・抽出し、抽出されたピッチを、ピッチパターンROM 33に記憶されたアクセント、フレーズ、音素の各パターンの加算によってモデル化する。また、ピッチ誤差抽出器 34により実音声のピッチとパターン化されたピッチの誤差を抽出し、各パターンとその誤差とを伝送することにより圧縮符号化した音声符号データを生成する。復号化器では、音声符号化データからアクセント・フレーズ・音素の各パターン、及びパターン化されたピッチと実音声のピッチの誤差を加算することによってピッチを生成・復号化する。



【特許請求の範囲】

【請求項1】 自然音声进行分析して所定区間のピッチパターンを抽出する音源抽出手段と、前記所定区間のピッチパターンモデルを複数記憶した記憶手段と、前記音源抽出手段によって抽出されたピッチパターンに最も良くマッチングするピッチパターンモデルを選択する選択手段と、前記選択されたピッチパターンモデルを用いて前記自然音声の所定区間のピッチパターンを符号化する手段とを備えることを特徴とする音声符号化装置。

【請求項2】 請求項1記載の音声符号化装置において、前記所定区間はアクセント成分単位の区間であることを特徴とする音声符号化装置。

【請求項3】 請求項1記載の音声符号化装置において、前記所定区間はアクセント成分単位の区間及びフレーズ成分単位の区間であることを特徴とする音声符号化装置。

【請求項4】 請求項1記載の音声符号化装置において、前記所定区間はアクセント成分単位の区間、フレーズ成分単位の区間及び音素成分単位の区間であることを特徴とする音声符号化装置。

【請求項5】 請求項1～4のいずれか1項記載の音声符号化装置において、実際のピッチパターンとピッチパターンモデルとの間のピッチ誤差を抽出するピッチ誤差抽出手段を備えることを特徴とする音声符号化装置。

【請求項6】 請求項5記載の音声符号化装置において、前記ピッチ誤差抽出手段によって抽出されたピッチ誤差を符号化するピッチ誤差時系列符号化手段を備えることを特徴とする音声符号化装置。

【請求項7】 請求項1～請求項6のいずれか1項記載の音声符号化装置において、前記符号化は前記ピッチパターンモデルに対応するコードによって行われることを特徴とする音声符号化装置。

【請求項8】 自然音声の所定区間のピッチパターンモデルを複数記憶した記憶手段と、前記ピッチパターンモデルに関連づけて符号化されたデータを前記ピッチパターンモデルを用いて復号化するピッチ生成手段とを備えることを特徴とする音声復号化装置。

【請求項9】 請求項8記載の音声復号化装置において、前記所定区間はアクセント成分単位の区間であり、前記ピッチ生成手段はアクセントパターン生成手段を備えることを特徴とする音声復号化装置。

【請求項10】 請求項8記載の音声復号化装置において、前記所定区間はアクセント成分単位の区間及びフレーズ成分単位の区間であり、前記ピッチ生成手段はアクセントパターン生成手段と、フレーズパターン生成手段と、前記アクセントパターン生成手段で生成されたアクセントパターンと前記フレーズパターン生成手段で生成されたフレーズパターンを加算する加算手段とを備えることを特徴とする音声復号化装置。

【請求項11】 請求項8記載の音声復号化装置におい

て、前記所定区間はアクセント成分単位の区間、フレーズ成分単位の区間及び音素成分単位の区間であり、前記ピッチ生成手段はアクセントパターン生成手段と、フレーズパターン生成手段と、音素パターン生成手段と、前記アクセントパターン生成手段で生成されたアクセントパターンと前記フレーズパターン生成手段で生成されたフレーズパターンと前記音素パターン生成手段で生成された音素パターンとを加算する加算手段とを備えることを特徴とする音声復号化装置。

10 【請求項12】 請求項8～11のいずれか1項記載の音声復号化装置において、前記データは自然音声のピッチパターンとピッチパターンモデルとのピッチ誤差を符号化した符号化データを含んでおり、前記ピッチ生成手段は前記ピッチ誤差を符号化したデータを復元するピッチ誤差復元手段を備え、前記ピッチ誤差復元手段で復元されたピッチ誤差が前記加算手段に入力されることを特徴とする音声復号化装置。

20 【請求項13】 請求項12記載の音声復号装置において、前記データは自然音声のピッチパターンとピッチパターンモデルとのピッチ誤差データを時系列的に符号化した符号化データを含んでおり、前記ピッチ生成手段は前記ピッチ誤差時系列符号化データを復元するピッチ誤差時系列復元手段を備え、前記ピッチ誤差時系列復元手段によって復元されたピッチ誤差データを前記ピッチ誤差復元手段に入力することを特徴とする音声復号化装置。

30 【請求項14】 請求項8～13のいずれか1項記載の音声復号化装置において、編集合成方式で音声を作成するとき、編集される音素素片のピッチが上限値及び下限値を越えないように、前記データに変更を加える制御手段を備えることを特徴とする音声復号化装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、入力音声进行分析、圧縮して符号化する音声符号化装置、及び符号化されたデータから音声を復号化し、合成する音声復号化装置に関する。

【0002】

40 【従来の技術】一般に、音声の符号化による音声符号化復号化装置では、その符号化の圧縮率が大きな問題となる。すなわち符号化音声を送信する場合、圧縮率が高いほど同じ伝送容量で多くの音声データを伝送可能である。また、符号化したデータを記憶装置に蓄積しておき、その蓄積された符号化データを基に音声を再生する場合でも、符号化の圧縮率が高いほど同じ記憶容量の記憶装置で多くの音声データを蓄積可能となる。

【0003】一般的な高効率な音声符号化では、音声の生成過程を考慮して、音声スペクトル特徴パラメータ、音源パラメータに分離して符号化する。その音源パラメータのうち、ピッチは音の高さを特徴付ける重要な

パラメータで、音声の符号化の際のパラメータとして一般的に広く用いられている。

【0004】音声のピッチ周波数のダイナミックレンジは、一般男性の場合、ゼロを含めて約50Hz～250Hz、一般女性の場合、ゼロを含めて約100Hz～400Hz程度である。このため、ピッチを符号化する場合、零を含めて約50～400のダイナミックレンジを持つ数値を圧縮符号化する必要がある。

【0005】従来、ピッチの圧縮符号化のためには、

(1) 人間の聴覚特性を利用して、ピッチ周波数を対数処理したものを量子化して圧縮したり、(2) ピッチの差分を符号化したりしていた。人間の音源の音声生成機構を考えた場合、ピッチは時間的に不連続にならない。このため、時系列にある時間間隔に音声の特徴パラメータを圧縮符号化する場合、分析した圧縮する時間間隔が十分に短い場合、その差分値のダイナミックレンジはピッチ周波数自体のダイナミックレンジに比べ小さくなる。前記(2)の方法は、このことを利用するものである。

【0006】また、記憶装置に符号化した音声データを蓄積して音声を再生する場合、記憶装置の蓄積容量を減少させる目的のため、発生する内容の一部分を別々に収録し、編集つなぎ合わせて合成音声を作る編集合成方式による音声合成方法が知られている。

【0007】

【発明が解決しようとする課題】現状のピッチの圧縮符号化は、聴覚特性に基づいて対数軸で量子化したり、差分の符号化によるダイナミックレンジの縮小化によって行っている。音声波形を分析し圧縮する時間間隔が十分に短い場合、そのピッチの差分値のダイナミックレンジはピッチ自体のダイナミックレンジに比べ小さくなる。このため、差分処理によってピッチの圧縮率を高めることが可能になる。しかしながら、分析する時間間隔を大きくした場合、その差分値のダイナミックレンジはあまり小さくならず、圧縮率を高くすることができないという問題がある。

【0008】そこで、本発明の目的は、このような従来技術の問題点を解決し、圧縮符号化の圧縮率をさらに高めることにある。また、記憶装置に符号化した音声データを蓄積編集して音声を編集合成する場合、収録した音素片間のピッチが連続的に接続されず、音の高さが急激に変わり、異音が発生したり、不自然に聞こえるという問題点があった。本発明の他の目的は、圧縮符号化された音素片を編集して合成音声を出力する編集合成方式において、ピッチの不連続性を解消することにある。

【0009】

【課題を解決するための手段】前記目的は、ピッチパターンをモデル化し、そのモデル化したパラメータで音声を圧縮符号化する、あるいは実際のピッチパターンとピッチパターンモデルの差の情報をさらに付加して圧縮符

号化することにより達成される。モデル化は、音声生成機構に立脚したモデルを用い、アクセント句単位、フレーズ単位、音素単位などの区間のピッチをまとめてモデル化を行うことでピッチ情報を圧縮する。

【0010】入力された音声のピッチを分析間隔単位に分析・抽出し、抽出されたピッチをアクセント、フレーズ、音素の各パターンの加算によってモデル化し、実音声のピッチとパターン化されたピッチの誤差と各パターンを伝送することにより圧縮符号化した音声符号データを生成する。そして、音声符号化データからアクセント・フレーズ・音素の各パターン、及びパターン化されたピッチと実音声のピッチの誤差を加算することによってピッチを生成・復号化する。

【0011】また、前記した音素間のピッチが不連続になる問題は、ピッチパターンをモデル化していないことに起因していると考えられる。したがって、ピッチ制御に音声生成機構に立脚したモデルを用いることにより、ピッチの不連続性をなくすことができる。このピッチの不連続性をなくすことによって、上述の音の高さが急激に変わり、異音が発生したり、不自然に聞こえるという問題が解消され、より自然性の高い音声合成を可能にすることができる。

【0012】さらに、モデル化圧縮されたピッチ情報のモデルパラメータをピッチモデルに基づいて制御変更するモデルパラメータの制御手段を備え、ピッチモデルに立脚しても、例えば大きなアクセント成分が連続した場合などピッチが大きくなりすぎないように、モデル化したパラメータを制御変更することにより、ピッチが高くなりすぎたり、低くなりすぎたりしないようにすることができる。

【0013】すなわち、本発明の音声符号化装置は、自然音声を分析して所定区間のピッチパターンを抽出する音源抽出手段と、所定区間のピッチパターンモデルを複数記憶した記憶手段と、音源抽出手段によって抽出されたピッチパターンに最も良くマッチングするピッチパターンモデルを選択する選択手段と、選択されたピッチパターンモデルを用いて自然音声の所定区間のピッチパターンを符号化する手段とを備えることを特徴とする。符号化はピッチパターンモデルに対応するコードによって行うことができる。

【0014】前記所定区間は、アクセント成分単位の区間、アクセント成分単位の区間及びフレーズ成分単位の区間、あるいはアクセント成分単位の区間、フレーズ成分単位の区間及び音素成分単位の区間とすることができる。また、実際のピッチパターンとピッチパターンモデルとの間のピッチ誤差を抽出するピッチ誤差抽出手段、あるいはピッチ誤差抽出手段によって抽出されたピッチ誤差を符号化するピッチ誤差時系列符号化手段を備えることもできる。

【0015】また、本発明による音声復号化装置は、自

然音声の所定区間のピッチパターンモデルを複数記憶した記憶手段と、ピッチパターンモデルに関連づけて符号化されたデータをピッチパターンモデルを用いて復号化するピッチ生成手段とを備えることを特徴とする。前記所定区間がアクセント成分単位の区間であるとき、ピッチ生成手段はアクセントパターン生成手段を備える。

【0016】所定区間がアクセント成分単位の区間及びフレーズ成分単位の区間であるとき、ピッチ生成手段はアクセントパターン生成手段と、フレーズパターン生成手段と、アクセントパターン生成手段で生成されたアクセントパターンとフレーズパターン生成手段で生成されたフレーズパターンを加算する加算手段とを備える。

【0017】所定区間がアクセント成分単位の区間、フレーズ成分単位の区間及び音素成分単位の区間であるとき、ピッチ生成手段はアクセントパターン生成手段と、フレーズパターン生成手段と、音素パターン生成手段と、アクセントパターン生成手段で生成されたアクセントパターンとフレーズパターン生成手段で生成されたフレーズパターンと音素パターン生成手段で生成された音素パターンとを加算する加算手段とを備える。

【0018】データが自然音声のピッチパターンとピッチパターンモデルとのピッチ誤差を符号化した符号化データを含むとき、ピッチ生成手段はピッチ誤差を符号化したデータを復元するピッチ誤差復元手段を備え、ピッチ誤差復元手段で復元されたピッチ誤差を加算手段によって加算する。

【0019】データが自然音声のピッチパターンとピッチパターンモデルとのピッチ誤差データを時系列的に符号化した符号化データを含んでいるとき、ピッチ生成手段はピッチ誤差時系列符号化データを復元するピッチ誤差時系列復元手段を備え、ピッチ誤差時系列復元手段によって復元されたピッチ誤差データをピッチ誤差復元手段に入力する。

【0020】また、前記音声復号化装置において、編集合成方式で音声合成するとき、編集される音素素片のピッチが上限値及び下限値を越えないように、前記データに変更を加える制御手段を備えることを特徴とする。

【0021】

【発明の実施の形態】以下、図面を参照して本発明の実施の形態を詳細に説明する。図1は、本発明による音声符号化復号化装置の構成を概略的に示したブロック図である。この音声符号化復号化装置は、音声圧縮符号化する音声圧縮符号化装置1と、圧縮符号化されたデータから音声を復号化して合成音声出力する音声復号化装置2からなる。

【0022】音声圧縮符号化装置1は、音声入力器20、スペクトル解析器21、音源抽出器22、ピッチモデル化器23、符号化器24から構成されている。また、音声復号化装置2は、符号化データ蓄積器（入力器）10、ピッチ生成器11、音源生成器12、制御器

13、スペクトル生成器14、音声合成出力器15から構成されている。

【0023】図2は音声圧縮符号化装置1中のピッチモデル化器23の詳細な構成例を示し、図3は音声復号化装置2中のピッチ生成器11の詳細な構成例を示す。ピッチモデル化器23は、図1に示すように、音源抽出器22及び符号化器24に接続され、図2に示すように、ピッチ周波数入力器30、パターンマッチング器31、ピッチパターン出力器32、ピッチパターンROM33、ピッチ誤差抽出器34等により構成されている。

【0024】またピッチ生成器11は、図1に示すように、符号化データ蓄積器（入力器）10及び音源生成器12に接続され、図3に示すように、ピッチパターン入力器40、アクセントパターン生成器41、フレーズパターン生成器42、音素パターン生成器43、加算器44、ピッチ誤差復号器45、ピッチ誤差時系列復元器46等から構成されている。

【0025】以下、上述の構成要素の動作を説明する。最初に、音声圧縮符号化装置1の動作について説明する。まず、音声入力器20は音声を入力して、分析フレーム単位に音声を切り出し、スペクトル解析器21に出力する。スペクトル解析器21は、音声入力器20から出力された音声のスペクトル分析を行い、声道情報と音源情報に分離し、声道情報を表すスペクトル情報を取り出し、符号化器24に出力すると共に、分離された音源情報を音源抽出器22に出力する。

【0026】音源抽出器22は、スペクトル解析器21で分離された音源情報を入力として、音声の音の高さを一般的に表すピッチ周波数（間隔）を抽出してピッチモデル化器23に出力する。また、ピッチ周波数以外の音源情報、パワー情報を符号化器24に出力する。ピッチモデル化器23は、音源抽出器22から抽出されたピッチ周波数をモデル化して情報量を削減・圧縮し、符号化器24にモデル化したパラメータを出力する。

【0027】ピッチモデル化器23で行うピッチのモデル化の1つとしては、日本語規則合成において規則化されよく用いられ、アクセント句の成分とフレーズの成分の重畳モデルによる藤崎モデル〔日本音響学会誌27, pp. 445-453, (1971)〕がある。この藤崎モデルに音素のパターンを追加したものとして、武田氏によるモデル〔「音素による変化を考慮した基本周波数パターン生成モデルと音声合成規則」、信学論(A) J73-A, pp. 379-386 (1990)〕がある。また数量化I類を用いたものとして阿部氏らによるモデル〔「基本周波数パターンの2階層制御方式」音講論集1-2-11, pp. 227-228 (1992, 3)〕がある。

【0028】このようなモデル化を行い音声認識に利用した例として、下平氏らによる韻律認識手法〔「ピッチパターンのクラスタリングによる連続音声の句境界検

出」、音講論集2-5-14, pp. 81-82 (1991)がある。これらの手法を応用することによりピッチモデル化器23で各アクセント句、フレーズ、音素のピッチパターンを抽出することができる。この際、ピッチのモデル化はアクセント句、フレーズなど音声現象や言語との対応を必ずしも正確に取る必要はなく、符号化効率の点から圧縮率が高くなるようにモデル化してもよい。

【0029】また、入力音声の内容が既知の場合、あるいは分析合成を同時に行う必要がなく、ROMなどに圧縮符号化データを蓄積しておけば良い場合等では、ピッチのモデル化にあたり自動抽出された各パターンの種類を入手をかけ、圧縮符号化データを修正し、高品位で自然で、かつ圧縮率の高いピッチ圧縮符号化データを作成することが可能である。

【0030】図2に示したピッチモデル化器23の構成例では、ピッチ周波数入力器30において、音源抽出器22で分析フレーム単位に抽出されたピッチ周波数を入力として、アクセント句、フレーズ、文単位などのまとまったピッチ周波数データ列をパターンマッチング器31へ出力する。ピッチパターンROM33は、予め音声のアクセント、フレーズ、音素のピッチパターンを調べ、クラスタリングなどを用いてパターンを登録しておく。ここで、ピッチパターンROM33は、アクセントパターンのみを持つことも、アクセントパターンとフレーズパターンを持つことも、全(アクセント、フレーズ、音素)パターンを持つこともできる。

【0031】パターンマッチング器31では、ピッチ周波数入力器22から入力されたピッチ周波数データ列とピッチパターンROM33に登録してあるピッチパターンとの比較を行い、最も良くマッチングしたパターンを入力されたピッチ周波数のパターンであると判定し、各アクセント、フレーズ、音素毎のパターンの種類をピッチパターン出力器32に出力する。

【0032】また、ピッチモデル化器23はピッチ誤差抽出器34を備えることもでき、その場合、ピッチ誤差抽出器34は、ピッチ周波数入力器30から出力されたピッチ周波数列とパターンマッチング器31でマッチングされたパターンの値列を入力としてピッチパターン出力器へその誤差(差分)値列を出力する。

【0033】さらに、ピッチ誤差抽出器35で抽出されたピッチ誤差を圧縮符号化するピッチ誤差時系列符号化器35を更に備えることもでき、その場合、ピッチ誤差時系列符号化器35は、ピッチ誤差抽出器34から出力された誤差値列を入力としてADPCM(Adaptive Differential Pulse Code Modulation)等で時系列符号化を行うことにより圧縮し、ピッチパターン出力器32へ圧縮された誤差データを出力する。

【0034】ピッチパターン出力器32は、パターンマッチング器31が出力した各アクセント、フレーズ、音

素毎のパターンを出力する。また、ピッチ誤差抽出器34を備えている場合には、ピッチ誤差抽出器34が出力する各パターンと実際のピッチ周波数列の誤差(差分)列をさらに出力する。ピッチ誤差時系列符号化器35を更に備えている場合には、ピッチパターン出力器32は、各パターンと実際のピッチ周波数列の誤差(差分)列の代わりに、ピッチ誤差時系列符号化器35が出力するADPCM等で時系列符号化を行ったパラメータを出力する。これら各アクセント、フレーズ、音素毎のパターン、各パターンと実際のピッチ周波数列の誤差(差分)列、ADPCM等で時系列符号化あるいはVQなどで符号化を行ったパラメータをモデル化ピッチ情報と呼ぶ。

【0035】最後に符号化器24は、スペクトル解析器21、音源抽出器22、ピッチモデル化器23で、各々分析・抽出されたスペクトル情報、音源情報・パワー情報、モデル化ピッチ情報を入力として、圧縮・符号化し、音声復号化装置2へ符号化データを出力する。

【0036】次に、音声復号化装置2について説明する。音声復号化装置2に設けられた符号化データ蓄積器(入力器)10は、音声圧縮符号化装置1と組み合わせる。音声を送信する場合は、伝送路からの入力器となり、音声符号化データとしてROM等に蓄積されたものを用いる場合には、符号化データの蓄積器となる。

【0037】符号化データ蓄積器(入力器)10は、フレーム単位に入力される圧縮符号化された符号化データを音源情報、スペクトル情報、パワー情報、モデル化ピッチ情報に分離し、モデル化ピッチ情報をピッチ生成器11に、音源情報及びパワー情報を音源生成器12に、スペクトル情報をスペクトル生成器14にそれぞれ出力する。ピッチ生成器11は、符号化データ蓄積器(入力器)10から出力されたモデル化ピッチ情報を入力とし、ピッチ周波数(間隔)を復号化し音源生成器12に出力する。

【0038】音声復号化装置2は、符号化した蓄積符号化音声データを編集して音声合成を行う(編集合成の場合、制御部13を備え、ピッチ周波数が高くなりすぎる場合、低くなりすぎる場合、ピッチ周波数が上限、下限を越えないようにアクセントパターン、フレーズパターンの大きさを制限するように、モデル化ピッチ情報の制御信号をピッチ生成器11に出力する。ピッチ生成器11は、制御部13からの制御信号を基にピッチ周波数が下限、上限を越えないようにモデル化ピッチ情報を変更し、ピッチ周波数(間隔)を生成して、音源生成器12に出力する。このことにより、ピッチが高すぎたり、低すぎて不自然に聞こえるという問題点を解消し、より自然性の高い音声合成が可能になる。

【0039】また編集合成により音声合成を行う場合以外でも、ピッチをモデル化することにより、音素片(音声素片)を接続して合成音声を再生する録音編集合成の

場合、ピッチが不連続に接続されることがなくなることで、ピッチが急激に変わり、異音が発生したり、不自然に聞こえるという問題点を解消し、より自然性の高い音声合成が可能になる。

【0040】音源生成器12は、符号化データ蓄積器（入力器）10から入力される音源情報、パワー情報と、ピッチ生成器11から出力されるピッチ周波数（間隔）を入力として音源を生成し、スペクトル生成器14に出力する。スペクトル生成器14は、符号化データ蓄積器（入力器）10から出力されたスペクトル情報に基づいて合成フィルタを構成し、音源生成器12で生成された音源にフィルタリングすることによって合成音声信号を生成する。合成音声出力器15は、スペクトル生成器14で生成された合成音声信号をD/A変換することによって合成音声を生成する。

【0041】図3はピッチ生成器11の構成の一例を示すものであり、図3を参照してピッチ生成器11について更に詳細に説明する。ピッチ生成器11のピッチパターン入力器40では、符号化データ蓄積器（入力器）10で分析フレーム単位に圧縮符号化されたモデル化ピッチ情報を入力とし、アクセントパターンをアクセントパターン生成器41へ、フレーズパターンをフレーズパターン生成器42へ、音素パターンを音素パターン生成器43へそれぞれ出力する。各パターンと実際のピッチ周波数列の誤差（差分）列のデータがある場合には、それをピッチ誤差復元器45へ出力し、ADPCM等で時系列符号化を行ったパラメータがある場合には、それをピッチ誤差時系列復元器46へそれぞれ出力する。

【0042】ここで、モデル化ピッチ情報がアクセントパターンのみを対象にする場合には、ピッチ生成器11は、ピッチパターン入力器40とアクセントパターン生成器41のみから構成される。アクセントパターン生成器41は、入力されたアクセントパターンからアクセント単位のピッチ周波数（間隔）列を生成、出力する。

【0043】モデル化ピッチ情報がアクセントパターン・フレーズパターンを対象にする場合には、ピッチ生成器11は、ピッチパターン入力器40とアクセントパターン生成器41、フレーズパターン生成器42、加算器44から構成される。フレーズパターン生成器42は、入力されたフレーズパターンからフレーズ単位のフレーズのピッチ周波数（間隔）列を生成、出力する。加算器44は、入力されるアクセント単位、フレーズ単位などのピッチ周波数（間隔）を加算して、元の音声のピッチ周波数（間隔）を出力する。

【0044】モデル化ピッチ情報が全（アクセント、フレーズ、音素）パターンを対象にする場合には、ピッチ生成器11は、ピッチパターン入力器40とアクセントパターン生成器41、フレーズパターン生成器42、音素パターン生成器43、加算器44から構成される。

【0045】モデル化ピッチ情報がモデル化したピッチ

と実際のピッチとの誤差（差分）列を含んでいる場合には、ピッチ生成器11は、ピッチ誤差復元器45も備える。ピッチ誤差復元器45は、ピッチパターン入力器40から出力された各パターンと実際のピッチ周波数列の誤差（差分）列を入力として加算器44へ転送する。

【0046】さらにモデル化ピッチ情報がADPCM等で時系列符号化を行ったパラメータを含む場合には、ピッチ誤差復元器45とともにピッチ誤差時系列復元器46をも備える。ピッチ誤差時系列復元器46は、ピッチパターン入力器40から出力されたADPCM等で各パターンの加算によるピッチ周波数と実際のピッチ周波数列の誤差（差分）列の時系列符号化を行ったパラメータを入力として、ADPCM等で各パターンの加算によるピッチ周波数と実際のピッチ周波数列の誤差（差分）列を推定、生成する。生成された誤差列はピッチ復元器45に出力される。

【0047】このように、ピッチを符号化・復号化する際に、ピッチをモデル化することにより、従来のピッチの符号化データに比較して少ないデータ量でピッチを符号化可能になり、少ない電送容量、少ないメモリ量で高温質の復号化音声を生成することができる。以下に、具体例をもとに本発明による処理の流れを説明する。4は、音声圧縮符号化装置1に音声入力として「次の電車は天理行きです。」を入力した場合の処理の一例を示している。

【0048】音源抽出器22では、入力された自然音声のピッチパターンを算出し、フレーズ、アクセント、音素の各成分に分解する。この成分への分解は、HHM（Hidden Markov Model）等の周知の音声認識による音素の自動切り出し、前述の下平氏らによる韻律認識手法〔「ピッチパタンのクラスタリングによる連続音声の句境界検出」、音講論集2-5-14, pp. 81-82 (1991)〕等の手法によって行うことができる。

【0049】ピッチモデル化器23では、各成分に対してピッチパターンROM33に予め登録してあるパターンのうち、最も一致するものをパターンマッチング器31により選択する。処理例としては、まずアクセントを切り出し、最もマッチングするアクセントパターン、フレーズパターンをROMに登録されたパターンから検索、抽出する。次に、個々の音素（子音）についての音素パターンをROMに登録された音素パターンから検索、抽出する。符号化器24では、検索された各パターンのコード番号を圧縮パターンとする。

【0050】前記パターンマッチングの際、時間軸の圧縮伸長率、ピッチパターンの開始値、終了値、最大値、最小値、圧縮伸張率の全部あるいは一部等のパラメータを用い、ピッチパターンROM33の登録パターンを変形する機能を持たせ、ROMパターンの削減を行い、各ピッチパターンのコード番号及び変形パラメータをそのパターンの圧縮パラメータとしてもよい。

11

【0051】ピッチパターンROM33に登録するパターンの一般的な作成方法としては、大量の自然音声の各ピッチパターンを抽出した後、クラスタリングを行い、ROMに入れるべきピッチパターンを決定する。この際、多くのピッチパターンをROM33に入れた場合、モデル化精度が高くなる反面、圧縮率の低下、処理量の増大を招く。

【0052】音声復号化装置2側のピッチ生成器11は、音声圧縮符号化装置1側で保持しているピッチパターンROMと同じピッチパターンを記憶したROMを持ち、符号化器24によって符号化されて送られてきたコード番号からピッチパターンを生成する。

【0053】

【発明の効果】以上より明らかなように、本発明では、ピッチをモデル化し、そのモデル化したパラメータで圧縮符号する、あるいは実際のピッチとピッチモデルの差をさらに付加し圧縮符号化する、あるいはさらに時系列的にそれらの差を推定することにより、音声符号化の際の圧縮率をさらに高めることが出来る。

【0054】さらに、本発明では、ピッチをモデル化することにより、音声を編集して合成音声出力する編集合成方式において、音声生成機構に立脚したモデルを用いることにより、接続する音声素片間にピッチが連続的

12

に接続されるため、より自然性の高い音声を合成することが出来る。

【図面の簡単な説明】

【図1】本発明による音声合成装置の一例の回路構成を示すブロック図。

【図2】ピッチモデル化器の構成例を示すブロック図。

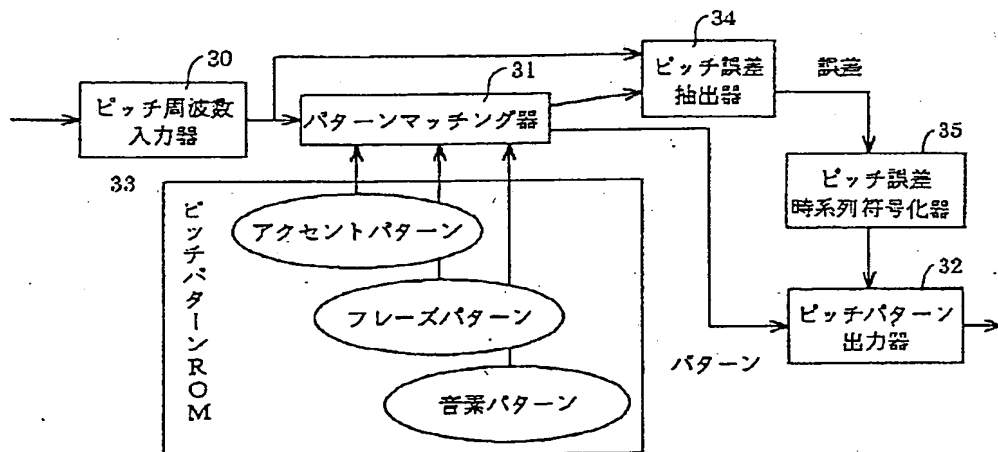
【図3】ピッチ生成器の構成例を示すブロック図。

【図4】自然音声からのパターン抽出の例を説明する図。

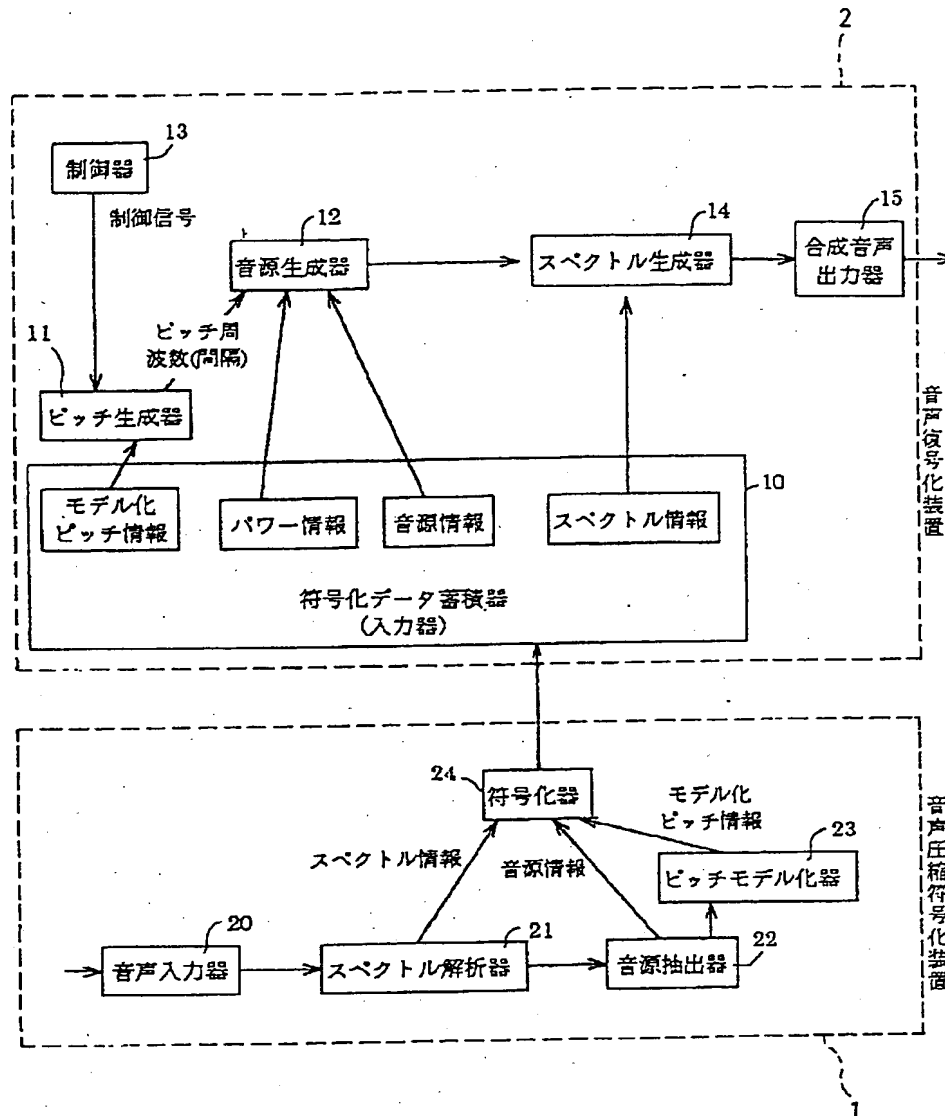
【符号の説明】

1…音声圧縮符号化装置、2…音声復号化装置、10…符号化データ蓄積器（入力器）、11…ピッチ生成器、12…音源生成器、13…制御機、14…スペクトル生成器、15…合成音声出力器、20…音声入力器、21…スペクトル解析器、22…音源抽出器、23…ピッチモデル化器、24…符号化器、30…ピッチ周波数入力器、31…パターンマッチング器、32…ピッチパターン出力器、33…ピッチパターンROM、34…ピッチ誤差抽出器、35…ピッチ誤差時系列符号化器、40…ピッチパターン入力器、41…アクセントパターン生成器、42…フレーズパターン生成器、43…音素パターン生成器、44…加算器、45…ピッチ誤差復元器、46…ピッチ誤差時系列復元器

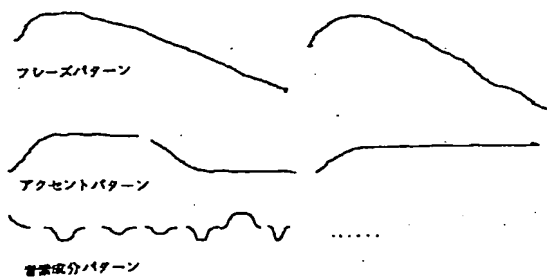
【図2】



【図1】



【図4】



【図3】

